Journal of Oral Rehabilitation 2014

Review

Diagnostic accuracy of portable instrumental devices to measure sleep bruxism: a systematic literature review of polysomnographic studies

D. MANFREDINI*, J. AHLBERG[†], T. CASTROFLORIO[‡], C. E. POGGIO[§], L. GUARDA-NARDINI* & F. LOBBEZOO[¶] *TMD Clinic, Department of Maxillofacial Surgery, University of Padova, Padova, Italy, [†]Department of Stomatognathic Physiology and Prosthetic Dentistry, Institute of Dentistry, University of Helsinki, Helsinki, Finland, [‡]Department of Surgical Sciences, Specialization School of Orthodontics, University of Turin, Turin, Italy, [§]Department of Prosthodontics, Eastman Institute for Oral Health, University of Rochester, Rochester, NY, USA and [¶]Department of Oral Kinesiology, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and VU University Amsterdam, MOVE Research Institute Amsterdam, Amsterdam, The Netherlands

SUMMARY This study systematically reviews the sleep bruxism (SB) literature published in the MEDLINE and Scopus databases to answer the following question: What is the validity of the different portable instrumental devices that have been proposed to measure SB if compared with polysomnographic (PSG) recordings assumed as the gold standard? Four clinical studies on humans, assessing the diagnostic accuracy of portable instrumental approaches (i.e. Bitestrip, electromyography (EMG)-telemetry recordings and Bruxoff) with respect to PSG, were included in the Methodological shortcomings identified by QUADAS-2 quality assessment. Findings showed contrasting results and supported only in part the validity of the described diagnostic devices with respect to PSG. The positive predictive value (PPV) of the Bitestrip device was 59-100%, with a sensitivity of 71-84.2%, whilst **EMG-telemetry** recordings

unacceptable rate of false-positive findings (76.9%), counterbalanced by an almost perfect sensitivity (98.8%). The Bruxoff device had the highest accuracy values, showing an excellent agreement with PSG for both manual (area under ROC = 0.98) and automatic scoring (0.96) options as well as for the simultaneous recording of events with respect to PSG (0.89-0.91). It can be concluded that the available information on the validity of portable instrumental diagnostic approaches with respect to PSG recordings is still scarce and not solid enough to support any non-PSG technique's employ as a stand-alone diagnostic method in the research setting, with the possible exception of the Bruxoff device that needs to be further confirmed with future investigations.

KEYWORDS: bruxism, sleep bruxism, diagnosis

Accepted for publication 8 June 2014

Introduction

Bruxism is an oral condition characterised by different activities of the jaw muscles (i.e. teeth clenching or grinding, mandible bracing or thrusting) and circadian manifestations (i.e. sleep or awake bruxism) (1). Over

the years, various bruxism topics have been addressed by means of systematic assessments of the published literature (2–7). Most reviews pointed out that the internal validity of the findings is often limited by methodological shortcomings related with the bruxism diagnosis (8, 9), because the wide majority of data came from studies adopting a self-reported bruxism detection. Such strategy is suitable, at best, to indicate a 'possible' bruxism (1).

A self-report approach is by far the most diffused in the dental literature, but it should be kept in mind that the proposed standards of reference for bruxism diagnosis require definite measurements of this phenomenon (1). Indeed, bruxism should be viewed as a continuous spectrum of muscle activities, whose physiopathology has yet to be fully clarified in its multifaceted neurological and musculoskeletal implications. Also, it must be remarked that definite criteria to measure bruxism during wakefulness are not available, despite the recent proposal to combine jaw muscles' electromyography (EMG) with an ecological momentary assessment (1). On the other hand, the most suitable approach to achieve sleep bruxism (SB) measurement, viz. polysomnography (PSG) with audio-video (AV) recordings (10, 11), has some limitations in its daily employ because of cost and feasibility reasons.

Bearing in mind these premises, some portable devices for measuring the EMG activity of jaw muscles in the home environment have been introduced in the research and clinical settings, in the attempt to partly solve the limitations related with the use of PSG (e.g. the high cost, its time-consuming nature, the complicated scoring criteria, the need for a sleep laboratory and the potential bias related with an unfamiliar sleeping environment). Notwithstanding that, an appraisal of the accuracy of the available portable measurement devices is lacking, especially related to the correlation with PSG findings. Given the absence of widely accepted standards for an awake bruxism diagnosis, this study will systematically review the literature on the diagnosis of SB in an attempt to find an answer to the following question: What is the validity of the different portable instrumental devices that have been proposed to measure SB if compared with PSG recordings assumed as the gold standard?

Materials and methods

Definition of diagnostic standard of reference for sleep bruxism

At present, widely accepted reference criteria for the diagnosis of SB refer to the original publication by

Lavigne et al. (10), who assessed the validity of PSG research diagnostic criteria for SB. In a case-control study on two 18-subject groups of bruxers and ageand sex-matched controls, the authors established sound cut-off thresholds for PSG-recorded jaw-motor activity to identify individuals with sleep bruxism, as diagnosed according to the American Sleep Disorders Association (ASDA) guidelines (12). To satisfy the ASDA criteria for SB, subjects had to grind their teeth during sleep for at least five times per week over the last 6 months (as referred by a bed partner) plus either exhibit tooth wear or shiny spots on restorations and/or morning masticatory muscle fatigue or pain and/or masseter muscle hypertrophy. The authors measured EMG activity of the right masseter and retained for analysis all EMG potentials with an amplitude >20% of the maximum voluntary contraction (MVC) level of the jaw-closing muscles. The resulting EMG bursts may combine to define different types of bruxism episodes (i.e. phasic/rhythmic, tonic or mixed), depending on the duration of each burst and of the between-burst intervals. SB could be detected with a sensitivity and specificity >80% when subjects had the following:

- 1 30 bruxism episodes per night or at least four episodes per hour of sleep; and
- **2** Six EMG bursts per bruxism episode and/or 25 EMG bursts per hour of sleep; and
- **3** At least two of the above episodes should be accompanied by AV-detected tooth grinding.

In a later publication, the same research group revised these criteria by retrospectively reassessing the sleep laboratory data recorded over the years (13). It was then suggested that the sensitivity and specificity of the proposed PSG criteria dropped down to less acceptable levels if the anamnestic criterion to confirm bruxism was reduced to a referral of tooth grinding for three (instead of five) nights per week. In the same study, it was proposed to identify a subthreshold group of moderate bruxers, showing more than two but <4 SB episodes per hour of sleep.

The 1996 PSG-based criteria and their successive 2007 update are commonly considered as the best available diagnostic method to detect SB. Therefore, they will be adopted here as the standard of reference for reviewing the validity of the other instrumental approaches proposed for SB measurement in a home environment.

Search strategy

On 9 April 2014, a systematic search in the medical literature was performed to identify all peer-reviewed English language papers potentially relevant to the review. The terms 'bruxism' and 'diagnosis' were introduced as search keywords in the two most qualified medical databases (i.e. National Library of Medicine's MEDLINE and Scopus) to retrieve lists of potential papers to be included in the review. The studies were selected for inclusion independently by two of the authors. The other authors contributed to the search expansion by checking for additional papers in the Google Scholar database and their own personal libraries and collections. All decisions on the definitive inclusion of a potentially relevant paper were taken by consensus.

The search allowed identifying 1293 and 552 citations in the MEDLINE and Scopus databases, respectively. Title and abstract ('tiab') screening was performed to select articles for full text retrieval.

The criteria for admittance in the systematic review were based on the type of study, because the inclusion was restricted to clinical studies on humans, assessing the diagnostic accuracy of any instrumental approach with respect to the above PSG criteria for SB diagnosis.

Quality assessment

In the attempt to increase the strength of this review, and in line with current needs to weigh the quality of the reviewed literature in systematic reviews, studies that were pertinent for inclusion underwent a quality assessment by adopting the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool (14). This is an improved, redesigned tool with respect to the original instrument (15). Although it is not specifically a reporting guideline, it represents a useful tool to rate the risk of bias and the applicability concerns of primary diagnostic studies included in systematic reviews. The tool comprises four key domains that discuss (i) patients selection; (ii) index test; (iii) reference standard, (iv) flow of patients through the study and timing of the index test and reference standard. For each domain, specific signalling questions are formulated, to help reviewers assessing each domain in terms of risk of bias; the first three domains are also assessed in terms of concerns regarding applicability.

Just as an example question, for the first domain (patient selection), the risk of bias should be evaluated by answering to the signalling questions 'Was a consecutive or random sample of patients enrolled?', 'Was a case-control design avoided?', 'Did the study avoid inappropriate conclusions?'. Reviewers are thus able to judge the risk of bias as 'low risk of bias', 'high risk of bias' or 'unclear'. The applicability rating for the same domain refers to the potential concerns that the included patients and setting do not match the review question. Reviewers are thus able to judge applicability as 'low applicability concerns', 'high applicability concerns' or 'unclear'. All the other signalling questions and specifications can be found in the original publication (14). In this investigation, QUADAS-2 ratings were assigned by two of the authors, who took each decision by consensus.

Results

As shown in Fig. 1, after excluding the citations that were clearly not pertinent for the review's aim, 12 papers were retrieved in full text and were assessed independently by two of the authors to reach consensus as to include/exclude the papers for/from systematic assessment. Consensus decision was to exclude eight of the 12 papers. Reasons for exclusion were that they were studies either adopting PSG as the unique approach to bruxism diagnosis (N = 3), including only convenience samples of pain patients (N = 2), describing the potential usefulness of other instrumental devices without comparing them with PSG (N = 2) or assessing the intra-examiner reliability of manual PSG/EMG scores (N = 1).

Two of the included papers adopted PSG criteria as the standard for comparison versus the Bitestrip device (16, 17), one of them dealt with a comparison of PSG criteria with EMG-telemetry recordings (18) and another one assessed the diagnostic accuracy of the Bruxoff device and the correlation of the findings with respect to PSG (19). All four studies were performed on single-night recordings with a portable PSG device, not allowing audio-visual recordings, and had the common aim of assessing the device's validity to measure the EMG activity of the masseter muscle by comparing it with PSG-based findings in either samples of bruxers or according to a case—control design. Moreover, the Bruxoff device also assessed

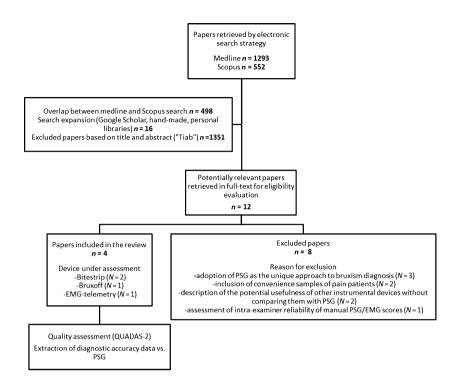


Fig. 1. Flow chart of the search strategy.

heart rate as part of the attempt to identify EMG episodes associated with an increase in heart rate (19). In three of four studies, the identified methodological problems were mainly related with the interpretation of the output of the device under assessment, which was produced at different thresholds with respect to the reference test (i.e. PSG) (16–18). The study on the Bruxoff device has potential bias related with the PSG-based SB diagnosis itself, as it was based on an EMG increase up to 10% MVC, and not 20%, as suggested in the PSG/SB criteria (19). All QUADAS-2 scores are shown in Table 1.

The extracted diagnostic accuracy data show contrasting results (Table 2). The positive predictive value (PPV) of the Bitestrip device was 59–100%, with a

sensitivity of 71–84·2% (16, 17), whilst EMG-telemetry recordings had an unacceptable rate of false-positive findings (76·9%), counterbalanced by an almost perfect sensitivity (98·8%) (18). The Bruxoff device had the highest accuracy values, showing an excellent agreement with PSG for both manual (area under ROC = 0·98) and automatic scoring options (0·96) as well as for the contemporaneity of events recorded with respect to PSG (0·89–0·91) (19). In the interpretation of the results, it must be pointed out that the PPV values, when calculated, were influenced by the prevalence of SB in the study samples. Also, importantly, the findings of three studies (16–18) appear to be strongly influenced by the fact that different EMG thresholds were used to diagnose sleep bruxism

Table 1. Quality assessment scores (QUADAS-2) of the four reviewed studies assessing the validity of the different approaches to sleep bruxism (SB) diagnosis if compared with polysomnographic (PSG) recordings assumed as the gold standard

	Risk of bias			Applicability concerns			
First author, year	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Castroflorio, 2014 (19)	Low	Unclear	Unclear	Low	Low	Low	Low
Mainieri, 2012 (17)	Low	High	Low	Low	Low	Low	Low
Yamaguchi, 2012 (18)	High	High	Low	Low	Low	Low	Low

Table 2. Summary of findings of the four reviewed studies assessing the validity of the different approaches to sleep bruxism (SB) diagnosis if compared with polysomnographic (PSG) recordings assumed as the gold standard. F, females; M, males; MVC, maximum voluntary clenching; ROC, receiver operating curve; PPV, positive predictive value; OSA, obstructive sleep apnoea

First author, year	Study sample	Diagnostic device under assessment	EMG cut-off threshold	PSG diagnosis	Findings (95% C.I.)
Castroflorio, 2014 (19)	N = 25 (12F, 13M, 28 ± 10·7 years) – case–control design (14 'probable' SB, 11 no SB)	Bruxoff device	10% MVC + 20% increase in heart rate	SB present or absent based on Lavigne <i>et al.</i> 's criteria (1996) (10) – 10% MVC	Manual scoring: Accuracy (ROC) = 89% Sensitivity = 83·3% Specificity = 84·6% Automatic scoring: Accuracy (ROC)=91% Sensitivity = 91·6% Specificity = 84·6%
Mainieri, 2012 (17)	$N=49$ (32F, 17M, $41\cdot 2\pm 12\cdot 9$ years) with a clinical history of SB	Bitestrip device	30% MVC	SB present or absent based on Lavigne <i>et al.</i> 's criteria (1996) (10) – 20% MVC	Agreement = 87.8% (75.8–94.3%) Kappa = 0.71 (0.44–0.97) Sensitivity = 84.2% (68.7–93.9%) PPV = 100% (89.1–100%)
Yamaguchi, 2012 (18)	N = 8 (50% F, $26.9 \pm 10.9 \text{ years})$ tooth grinders	EMG-telemetry	Two times higher than baseline	SB present or absent based on Lavigne <i>et al.</i> 's criteria (1996) (10) - 20% MVC	Sensitivity = 98% PPV = $23 \cdot 1\%$
Shochat, 2007 (16)	N = 18 (13M, 5F, 31 ± 13 years) – case–control design (6 SB, 4 OSA, 8 non-patient)	Bitestrip device	30% MVC	SB present or absent based on Lavigne <i>et al.</i> 's criteria (1996) (10) – 20% MVC	Sensitivity = 71–72% PPV = 59–81%

events (i.e. 30% MVC for the Bitestrip device; and EMG amplitude two times higher than rest values for the EMG-telemetry recordings) with respect to the thresholds adopted for PSG diagnosis (i.e. 20% MVC), so that a direct comparison with PSG criteria was actually impossible to perform. As for the Bruxoff study (19), the EMG threshold was set at 10% MVC for both the test and reference device (19).

Discussion

The literature on several bruxism topics pointed out inconsistencies due to the different strategies and definitions adopted for diagnostic purposes, which seem to influence the findings of many studies, so that a need emerged to carefully appraise the knowledge on the diagnosis of bruxism (4, 7–9). Thus, the present systematic review was specifically designed to answer a research question in the field of SB diagnosis.

The review was aimed at the assessment of the validity of the different portable instrumental

approaches to SB diagnosis if compared with PSG recordings assumed as the gold standard. Only four papers were found that satisfied the criteria for inclusion in the review, and in general, they did not allow to retrieve conclusive information. In particular, several concerns with respect to the studies' risk of bias in the interpretation of the index test were identified. Indeed, despite the potential limitations related with the dichotomic (i.e. low versus high [risk of bias and applicability concerns]) quality assessment approach here adopted, it seems reasonable to suggest that the accuracy values for the SB diagnosis were strongly biased by the different levels of EMG activity that were adopted as cut-off levels in the index (viz. test) and reference (viz. PSG) instruments (16-19). The adoption of such different strategies may be in part justified by the different types of EMG equipment, using different band-pass and notch filtering.

Not all devices showed an optimal validity to measure what they are supposed to measure, viz. oromotor activity during sleep associated with SB, if PSG

As a suggestion for future studies attempting to get deeper into the definition of an 'ideal' SB diagnosis, some considerations are worthy to be made on the issue of grinding sounds during sleep. Indeed, the need to include grinding sounds to diagnose SB by means of the reference AV/PSG criteria, which require that at least two of the SB episodes per hour should be accompanied by AV-detected tooth grinding, may exclude some other motor phenomena (e.g. clenching activities and mandible thrusting) from the

test and reference devices (19).

SB bruxism diagnosis that may be more important in the pathogenesis of pain in the jaw muscles and temporomandibular joint (TMJ) (21). Thus, independently from the accompanying grinding sounds, future studies on SB diagnosis should differentiate the different patterns of EMG activity as to put them into correlation with their potential clinical consequences. To do that, it is important to assess the correlation of PSG-based diagnoses with findings based on clinically diagnosed and self-reported bruxism, as recently carried out in a large-sample epidemiological paper (22). Notwithstanding that, based on the often contrasting findings between PSG and questionnairebased studies in the field of the TMD-bruxism literature (7), it cannot be excluded that other approaches based on patients' history and clinical assessment may be even more useful than PSG to diagnose clinically relevant bruxism in TMD patients (23). Indeed, it cannot be disregarded that most PSG studies were performed in super-selected samples of subjects without any other medical problems or environmental risk factors (e.g. other sleep disorders, psychological impairment, alcohol or caffeine abuse, heavy smoking) that in the clinical setting account for most part of the subjects with self-reported bruxism (24). Notwithstanding all the above issues that are yet to be clarified before reaching an ideal diagnosis of SB, especially as far as its clinical implications are concerned, this review's findings support the need for adopting PSG recordings, and not only EMG alone, to achieve a sound diagnosis of SB according to the current reference criteria. Keeping this in mind, the device adopting a combined EMG and electrocardiographic (ECG) recordings showed an increased accuracy with respect to the EMG-based devices and may represent a promising, simple tool for the diagnosis of SB.

Conclusions

Based on this systematic assessment of the literature on SB diagnosis, it can be concluded that the available information on the validity of portable instrumental diagnostic approaches with respect to PSG recordings is not solid enough to support any non-PSG technique's employ as a stand-alone diagnostic method in the research setting, with the possible exception of the Bruxoff device that needs to be further confirmed with future investigations.

Disclosures

No ethical approval was needed for this manuscript. The authors declare they received no funding for this investigation. The authors declare J.A., C.E.P., L.G.N. and F.L. have no conflicts of interest with this manuscript. T.C. and D.M. were involved in the research project on the Bruxoff device (ref#19), but none of them was in any financial relationships with the manufacturer of the device or has any other potential conflicts of interest.

References

- Lobbezoo F, Ahlberg J, Glaros A, Kato T, Koyano K, Lavigne GJ et al. Bruxism defined and graded: an international consensus. J Oral Rehabil. 2013;40:2–4.
- Lobbezoo F, Naeije M. Bruxism is mainly regulated centrally, not peripherally. J Oral Rehabil. 2001;28:1085–1091.
- Kato T, Thie NM, Huynh N, Miyawaki S, Lavigne GJ. Topical review: sleep bruxism and the role of peripheral sensory influences. J Orofac Pain. 2003;17:191–213.
- 4. Manfredini D, Lobbezoo F. Role of psychosocial factors in the etiology of bruxism. J Orofac Pain. 2009;23:153–166.
- Lobbezoo F, Ahlberg J, Manfredini D, Winocur E. Are bruxism and the bite causally related? J Oral Rehabil. 2012;39:489–501.
- Lobbezoo F, Brouwers JE, Cune MS, Naeije M. Dental implants in patients with bruxing habits. J Oral Rehabil. 2006;33:152–159.
- Manfredini D, Lobbezoo F. Relationship between bruxism and temporomandibular disorders: a systematic review of literature from 1998 to 2008. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2010;109:e26–e50.
- Manfredini D, Winocur E, Guarda-Nardini L, Paesani D, Lobbezoo F. Epidemiology of bruxism in adults. A systematic review of literature. J Orofac Pain. 2013;27:99–110.
- 9. Koyano K, Tsukiyama Y, Ichiki R, Kuwata T. Assessment of bruxism in the clinic. J Oral Rehabil. 2008;35:495–508.
- Lavigne GJ, Romprè PH, Montplaisir JY. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. J Dent Res. 1996;75:546–552.
- 11. Ikeda T, Nishigawa K, Kondo K, Takeuchi H, Clark GT. Criteria for the detection of sleep-associated bruxism in humans. J Orofac Pain. 1996;10:270–282.
- Thorpy MJ. ICSD International Classification of Sleep Disorders: Diagnostic and Coding Manual. Rochester, MN: Diagnostic Classification Steering Committee, American Sleep Disorders Association; 1990.
- 13. Romprè PH, Daigle-Landry D, Guitard F, Montplaisir JY, Lavigne GJ. Identification of a sleep bruxism subgroup with a higher risk of pain. J Dent Res. 2007;86:837–842.

- 14. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–536.
- 15. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25.
- 16. Shochat T, Gavish A, Arons E, Hadas N, Molotsky A, Lavie P *et al.* Validation of the BiteStrip screener for sleep bruxism. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2007;104:e32–e39.
- 17. Mainieri VC, Saueressig AC, Pattussi MP, Fagondes SC, Grossi ML. Validation of the Bitestrip versus polysomnography in the diagnosis of patients with a clinical history of sleep bruxism. Oral Surg Oral Med Oral Pathol Oral Radiol. 2012;113:612–617.
- Yamaguchi T, Abe S, Rompré PH, Manzini C, Lavigne GJ. Comparison of ambulatory and polysomnographic recording of jaw muscle activity during sleep in normal subjects. J Oral Rehabil. 2012;39:2–10.
- Castroflorio T, Deregibus A, Bargellini A, Debernardi C, Manfredini D. Detection of sleep bruxism: comparison between an electromyographic and electrocardiographic portable holter and polysomnography. J Oral Rehabil. 2014;41:163–169.
- Carra MC, Huynh N, Lavigne GJ. Diagnostic accuracy of sleep bruxism scoring in absence of audio-video recording: a pilot study. Sleep Breath. 2014. May 3 [Epub Ahead of Print].
- Manfredini D, Vano M, Peretta R, Guarda-Nardini L. Jaw clenching effects in relation to two extreme occlusal features: patterns of diagnoses in a TMD patient population. Cranio. 2014;32:45–50.
- 22. Maluly M, Andersen ML, Dal-Fabbro C, Garbuio S, Bittencourt L, de Siqueira JTT *et al.* Poylsomnographic study of the prevalence of sleep bruxism in a population sample. J Dent Res. 2013;92(7 Suppl):97S–103S.
- Paesani DA, Lobbezoo F, Gelos C, Guarda-Nardini L, Ahlberg J, Manfredini D. Correlation between self-reported and clinically based diagnoses of bruxism in temporomandibular disorders patients. J Oral Rehabil. 2013;40:803–809.
- 24. Ahlberg J, Lobbezoo F, Ahlberg K, Manfredini D, Hublin C, Sinisalo J *et al.* Self-reported bruxism mirrors anxiety and stress in adults. Med Oral Patol Oral Cir Buc. 2013;18:e7–e11.

Correspondence: Daniele Manfredini, Viale XX Settembre 298, 54033 Marina di Carrara (MS), Italy.
E-mail: daniele.manfredini@tin.it